상품 리뷰 요약에서의 문맥 정보를 이용한 의견 분류 방법

(A Sentiment Classification Method Using Context Information in Product Review Summarization)

양 정 연 † 명 재 석 † 이 상 구 ††
(Jung-yeon Yang) (Jaeseok Myung) (Sang-goo Lee)

요 약 e비즈니스가 활발히 이루어지면서 소비자들은 온라인 쇼핑몰을 통해 수많은 상품을 접할 수 있게 되었고, 상품구매 시 다른 사람들의 리뷰를 참고하게 되었다. 하지만, 리뷰의 수도 많아짐에 따라 소비자가 모든 리뷰들을 살펴보기가 힘들다는 문제점이 대두되었으며 이를 해결하기 위해서 리뷰의 상품에 대한 평가를 요약하고 성향을 파악하는 오피니언 마이닝 연구가 나타나게 되었다. 본 논문에서는 상품리뷰를 대상으로 오피니언 마이닝을 수행하는 경우 어휘의 의견 성향을 파악할 때, 문맥정보를 활용하여 기존의 의견분류방법 보다 좀 더 정확한 의견 판단이 가능한 방법에 대해 다루고 있다. 이를 위해, 어휘가 사용될 때의 문맥정보를 정의하고 이를 의견분류에 적용하는 방법을 제안하였으며, 실험을 통하여 기존 연구보다 상황별 알맞은 의견분류가 가능함을 보였다. 또한 수작업으로 말뭉치의 핵심 어휘들을 정의했던 기존 연구들에서의 방식에서 벗어나, 리뷰본문과 리뷰점수를 활용하여 자동으로 상황에 맞는 말뭉치를 구축하는 방법도 제안하였다. 이를 통해 상품리뷰에 나타난 어휘들의 문맥에 맞는 의미 성향을 정확하고 쉽게 판별해 낼 수 있게 되었다.

키워드 : 오피니언 마이닝, 의견분류, 상품리뷰요약, 문맥정보, e비즈니스, 상품정보

Abstract As the trend of e-business activities develop, customers come into contact with products through on-line shopping sites and lots of customers refer product reviews before the purchasing on-line. However, as the volume of product reviews grow, it takes a great deal of time and effort for customers to read and evaluate voluminous product reviews. Lately, attention is being paid to Opinion Mining(OM) as one of the effective solutions to this problem. In this paper, we propose an efficient method for opinion sentiment classification of product reviews using product specific context information of words occurred in the reviews. We define the context information of words and propose the application of context for sentiment classification and we show the performance of our method through the experiments. Additionally, in case of word corpus construction, we propose the method to construct word corpus automatically using the review texts and review scores in order to prevent traditional manual process. In consequence, we can easily get exact sentiment polarities of opinion words in product reviews.

Key words : Opinion Mining, Sentiment Classification, Product review summarization, Context information, e-business, Product Information

· 본 연구는 지식경제부 및정보통신연구진흥원의대학 IT연구센터 육성·지원사업(IITA-2008-C1090-0801-0031)의 연구결과로 수행되었음

† 학생회원 : 서울대학교 컴퓨터공학부
jyyang@europa.snu.ac.kr
jsmyung@europa.snu.ac.kr
†† 종신회원 : 서울대학교 컴퓨터공학부 교수
sglee@europa.snu.ac.kr

논문접수 : 2009년 1월 14일
심사완료 : 2009년 5월 21일

Copyright©2009 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 데이타베이스 제36권 제4호(2009.8)

1. 서 론

인터넷을 통한 전자상거래 활동이 커질수록 상품리뷰의 중요성은 점점 더 커지고 있다. 하지만, 상품리뷰 또한 그 양이 많아서 소비자들이 쉽게 모든 내용을 판단하기는 어렵다는 문제점을 가지고 있다. 이러한 문제점을 해결하기 위해 상품리뷰와 같은 콘텐츠의 의미적 정보를 파악하고자 하는 오피니언 마이닝 연구 분야가 나타나게 되었다. 오피니언 마이닝은 텍스트 마이닝의 한 분야로써, 특히 콘텐츠의 저자가 말하고자 하는 내용을 분석하여 요약하는 것이 목적이라고 할 수 있으며, 영화리뷰, 상품리뷰 등의 콘텐츠들을 대상으로 이루어진다. 오피니언 마이닝의 경우, 문서의 전체적인 의미를 파악하는 것이라는 점에서 그 성격과 접근 방향이 기존의 텍스트 마이닝과는 약간은 다르며 대표적으로 상품리뷰 요약에 활용되고 있다. 예를 들어 어떤 사용자가 가정용 냉장고를 사용하고 작성한 리뷰문서에 냉장고를 설치하고 나서, 이번 달 전기요금이 두 배나 더 나온 것을 보니, 소비전력 효율성 아주 안 좋은 것 같아요."라는 문장이 있을 때, 냉장고의 소비전력효율성이 나쁘다는 의견정보를 뽑아내는데 활용되는 기술이 오피니언 마이닝 기술이다.

오피니언 마이닝은 일반적으로 크게 특징추출, 의견분석, 요약 및 표현 과정으로 나뉘는데[1-3], 본 논문에서는 의미를 판단하는 주요단계인 의견분석 중 의견분류에 대한 방법을 다루고 있다. 의견분류는 특정 어휘가 긍정적 또는 부정적으로 사용된 것을 판단하는 작업으로 오피니언 마이닝의 목적을 수행하기 위한 중요한 과정이다. 이와 관련하여 많은 연구들이 수행되었는데, 대표적인 방법이 어휘들의 상호출현빈도를 활용한 PMI값을 계산하여 의견 성향을 판별하는 방식이다. 이는 전문가가 긍정적 또는 부정적인 의미의 핵심어휘들을 미리 정의한 후에 인터넷에 존재하는 문서들로부터 어휘들의 출현빈도를 계산해 내어, 특정 어휘의 긍정적 또는 부정적 성향을 판단하는 방식이다[4,5].

오피니언 마이닝 기술을 활용하여 온라인 쇼핑몰의 상품리뷰를 요약하는 방법들에 대한 기존의 연구들에서는 크게 2가지의 단점이 관찰된다. 첫째, 오피니언 마이닝은 문서의 의미정보를 파악하는 것이 목적으로, 어휘들이 어떤 상황에서 무슨 의미로 사용되어졌는지가 중요하다. 하지만 고정된 의미적 극성을 갖는 어휘의 활용이 중심이 된 기존방법의 경우 문맥에 따라 다른 의미적 극성을 가질 수 있는 어휘들의 문서 내에서의 의미를 파악하기에는 어려운 경우가 존재한다. 둘째, 분석하려는 콘텐츠별로 전문가가 핵심 어휘들을 미리 정의해야만 하는 것이다. 이는 수작업으로 이루어지며, 대상 콘텐츠의 성격 및 도메인이 바뀔 때마다 다시 정의해야 하는 비효율성을 지니고 있다.

본 연구에서는 온라인 쇼핑몰에서 제공하는 상품리뷰를 요약할 때, 전문가의 수작업에 의존하며, 광범위하고 고정적인 의미적 극성을 가진 어휘의 활용에 기반하고 있는 기존방법의 문제점을 해결하고, 도메인별로 정확한 어휘의 의견분류를 수행하기 위한 방법에 대하여 다루고 있다. 특정 도메인 및 상황에 맞는 어휘의 의미를 파악하기 위하여 어휘가 사용된 문맥정보를 고려하였으며, 의견 성향 판별을 위해 사용되는 긍정적, 부정적 말뭉치의 구축을 자동으로 수행하여 분류작업의 유연성을 높였다.

2. 관련연구

2.1 오피니언 마이닝

오피니언 마이닝은 리뷰데이터와 같은 대량의 정보 속에서 유용한 정보를 찾아내는 것이라는 특징을 가지고 있다. 이러한 오피니언 마이닝은 최근에 들어 활발히 연구되어 왔으며, 그 기반이 되는 기술은 자연어처리, 텍스트 마이닝, 통계 등의 분야로부터 기원하였다. 오피니언 마이닝은 대체로 크게 특징 추출, 의견분류, 요약 및 표현 등의 3가지 단계로 이루어진다. 첫째, 특징 추출 단계는 유용한 정보라고 판단되는 여러 특징들을 정의하고 추출해 내는 단계이다. 이때 단순히 특징만을 추출하는 것이 아니라 해당 특징이 어떤 의미를 가지는가에 대한 의견을 나타내는 어휘정보도 함께 추출된다. 둘째, 의견분류 단계에서는 추출된 특징과 의견을 나타내는 어휘가 해당 정보소스에서 어떤 의미로 사용되었는가에 대한 판단 및 분류를 하는 단계이다. 셋째, 요약 및 표현단계에서는 의견 성향이 밝혀진 의견정보들을 요약하여 전체 정보의 내용을 효율적으로 사용자에게 전달하는 단계이다. 이와 같은 오피니언 마이닝의 단계별 수행을 위한 다양한 방식들이 연구되고 있으며, 자연어처리기술 기반에서 통계적 기법에 이르기까지 여러 분야의 기술이 접목되고 있다[6-9].

초반의 연구들은 자연어처리기술을 기반으로 하여 활발히 수행되어져 왔으며, 최근에는 자연어처리기반 방식의 실제 적용에 대한 어려움을 해결하고자 상품리뷰요약 분야에서 통계적 확률론에 기초한 연구들이 다수 수행되어져 왔고, 기존 자연어처리기술과 통합되어 서로의 장점만을 추구하고자 한 연구 방법도 있었다[10-12]. 또한, 상품 전체에 대한 평가에서 벗어나 상품이 가지는 개별적인 특징들을 기준으로 전체 리뷰를 요약하는 방법의 연구도 수행되었다[13].

본 논문에서는 위와 같은 오피니언 마이닝의 단계 중에서 의견분류에 관한 효율적인 방법에 대하여 논하고

있다. 단순한 정보의 요약이 아닌 의미를 파악한다는 것이 오피니언 마이닝의 특징임을 감안했을 때, 의견정보의 올바른 의견분류는 반드시 필요한 과정이다.

2.2 Point-wise Mutual Information (PMI)

의견분류를 수행하기 위한 방법 중, 많이 연구되고 가장 널리 사용되는 개념으로 PMI가 있다. PMI는 두 단어가 서로 얼마나 밀접한 관계를 가지고 있는가를 확률모델 기반으로 계산하는 방법이며 다음과 같이 표현된다.

$$PMI(word1, word2) = \log_2 \frac{p(word1, word2)}{p(word1)p(word2)} \quad (1)$$

위의 식에서 볼 수 있듯이 PMI는 각 어휘의 단일 발생 확률 및 동시발생 확률 값을 활용한 확률기반 값으로써 계산되어진다. 즉, $p(word1)$, $p(word2)$는 전체 상품리뷰문서 중에서 각 어휘가 나타난 상품리뷰문서의 확률을 의미하며, $p(word1, word2)$은 전체 상품리뷰문서 중 두 어휘가 함께 나타난 상품리뷰문서의 확률을 의미한다. 위의 식을 통해 구해진 PMI값이 0인 경우, 두 단어 사이의 관계는 완전히 독립적인 것으로 볼 수 있으며, 반대로 두 단어 사이에 어떤 연관성이 존재하는 경우 0이 아닌 값으로 그 연관성의 정도가 표현된다. 두 단어의 관계가 긍정적이라면 양수의 값을 가지며, 서로 반대의 성향을 가진 관계라면 음의 값을 갖는다. 많은 오피니언 마이닝 연구에서는 이러한 두 단어의 연관성을 나타내는 PMI 값을 이용하여 의견정보의 의견분류를 수행한다.

PMI 값을 활용한 대표적인 연구로는 웹문서를 대상으로 특정어휘를 검색하였을 때, 검색된 문서의 수를 활용한 방법이 있다. 이 방법에서는 우선 긍정적인 의미의 어휘집합과 부정적인 의미의 어휘집합을 각각 전문가가 미리 정의를 내린다. 각 어휘집합이 정의되면 그 안의 개별 어휘들과 의미성향을 알고자 하는 특정 어휘사이의 PMI값을 검색엔진을 이용하여 계산한다.

$$SO-PMI(word) \quad (2)$$

$$= \sum_{pword \in Pwords} PMI(word, pword) - \sum_{nword \in Nwords} PMI(word, nword)$$

$$PMI(word1, word2) \quad (3)$$

$$= \log_2 \left(\frac{\frac{1}{N} hits(word1\ NEAR\ word2)}{\frac{1}{N} hits(word1) \frac{1}{N} hits(word2)} \right)$$

이 때, (2)와 같이 긍정적 어휘집합과의 PMI값과 부정적 어휘집합과의 PMI값의 차를 이용하여 특정 어휘의 의미적 성향을 계산하게 된다. 이 때, Pwords와 Nwords는 전문가가 사전에 정의한 긍정적, 부정적 의미의 어휘집합을 의미한다. 즉, 의미극성을 알고자 하는 어휘와 사전 정의된 어휘집합사이의 상관관계를 활용하여 최종 의미극성을 구한다. 식 (3)과 같이 각 어휘와 긍정적 또는 부정적 어휘와의 상관관계는 대상이 되는 두 어휘를 웹문서 대상으로 'near'연산 검색을 한 결과로 얻어진 웹문서의 수를 이용하여 구하게 되며, 사전 정의된 여러 어휘집합들에 대한 PMI값이 구해지면 식 (2)와 같이 긍정적인 경우와 부정적인 경우의 차이값으로 최종 의미극성을 판별한다[3,4].

3. 문맥정보를 이용한 의견분류 기법

3.1 문맥정보 정의 및 필요성

오피니언 마이닝을 통한 상품리뷰요약에서 특정 어휘의 문맥적 의미를 정확하게 파악하기 위해서는 어휘가 쓰인 상황적 정보, 즉 문맥(Context)을 반드시 고려해야만 한다. 무엇보다 어휘가 어떤 의미로써 사용되어져서 하나의 의견을 형성하고 있는지가 중요한 문제이므로, 의견을 나타내는 어휘자체만의 의미가 아닌, 주변 문맥적 상황을 모두 고려한 의견분류가 되어야 한다. 본 논문에서는 상품리뷰의 오피니언 마이닝에 있어서 어휘의 문맥정보를 다음과 같이 정의한다.

문맥정보 = (상품분류, 상품특징, 표현 어휘, 리뷰점수)

상품리뷰에서 의견은 어떤 상품분류, 어떤 특징에 대하여 어떤 어휘로 표현하고 있느냐에 따라 그 의견의 의미적 극성이 결정된다. 예를 들어, '작다'라는 표현의 어휘가 '업소용 냉장고'라는 상품분류의 '용량'이라는 특징에 사용된 경우 부정적인 의미를 갖는다. 하지만, 똑같은 '작다'라는 어휘가 '휴대용 음악기기'상품의 '크기' 특징에 사용된 경우에는 완전히 반대의 의미를 갖게 된다. 하지만 이러한 상황정보에 대한 고려가 없이, 단순히 '작다'라는 어휘의 일반적인 의견 성향을 파악한다면 그 의미가 매우 모호하게 된다. 이와 같이 상품리뷰 요약에서 어휘의 의견분류 시에는 반드시 상품분류와 특징이 함께 고려되어야 하는 것이다. 또한 의견이 서술된 리뷰에서 리뷰어가 상품에 대해 부여한 리뷰점수도 의견의 문맥정보에 해당한다. 리뷰점수는 별점과 같은 형태로 상품단위로 주어지는 값이며, 특정 어휘가 높은 점수를 받은 리뷰에서 나타났는지 혹은 낮은 점수를 받은 리뷰에서 나타났는지에 따라 그 의미적극성에 영향을 준다. 더욱이 리뷰점수는 리뷰어가 부여한 컴퓨터가 이해할 수 있는 명확히 표현된 평가값이라는 점에서 그 활용의 중요성이 크다.

기존의 오피니언 마이닝 연구들에서의 의견분류 방법에서는 어휘의 의미를 파악하는데 있어서 이러한 문맥정보를 고려하지 못하였다. 워드넷[14] 및 웹문서 등의 일반적인 말뭉치에서의 어휘간 관계정보를 활용하는 경

우, 이는 풍부한 어휘사전이 될 수 있지만, 해당 어휘가 정확히 어떤 상황 및 어떤 의미로 사용되었는지에 대한 판단을 하기에는 어휘데이터가 너무나 일반적 활용에 근거하고 있다는 문제점을 가지고 있다.

본 연구에서는 위와 같은 연구방식의 문제점이 존재함을 ePinions.com의 실제 상품리뷰데이터를 분석함으로써 밝힐 수 있었다. 휴대전화 상품분류와 디지털카메라 상품분류를 대상으로 분석해본 결과 <상품특징, 표현어휘>로 이루어진 의견정보 중에서 약 14~17%의 의견정보가 표현어휘 만의 의미로는 정확한 의미판단이 되지 않는 것으로 관찰되었다. 이와 같은 상황에서 어휘의 상황적 의미를 정확하게 파악하기 위해서는 본 연구에서 제안하는 어휘의 문맥적 상황정보를 고려한 의견분류 방법이 필요한 것이다.

3.2 문맥정보의 반영

본 절에서는 상품리뷰로부터 어휘가 사용된 문맥정보를 얻기 위한 방법 및 이를 의견분류에 적용하는 방법에 대하여 서술하였다. 우선 각 표현어휘의 문맥정보는 의견분류에 사용되는 말뭉치사전의 수집대상을 특정 상품군의 리뷰데이터로 제한함으로써 얻을 수 있다. 즉, 말뭉치사전을 워드넷이나 웹문서와 같은 일반적 의미의 광범위한 말뭉치사전이 아닌 해당 상품군에 대한 어휘들로 이루어진 말뭉치들로 구성하는 것이다.

다음으로, 상품특징에 대한 상황정보는 의견을 나타내는 어휘와 쌍을 이루어 처리함으로써 얻을 수 있다. 의미를 표현하는 어휘들의 정보만을 이용하여 그 의미적 성향을 추정하는 방식에서 벗어나, 모든 어휘의 의미를 추정할 때 해당 어휘가 수식하는 특징들을 함께 고려해야하는 것이다. <상품분류, 상품특징, 표현어휘, 리뷰점수>의 네 가지 차원의 문맥적 상황정보를 하나의 단위로 처리하는 것이 본 연구의 차이점이며, 이로 인하여 정확한 상황별 의견을 판별할 수 있는 것이다.

앞에서 설명한 특징을 나타내는 어휘와 의견을 표현하는 어휘 사이의 관계정보를 활용하기 위하여 본 연구에서는 PMI값을 두 어휘에 적용하였으며, 상품정보를 적용한 말뭉치사전을 활용하여 해당 의견의 최종 의견 성향을 계산하였다. 이 값을 구하는 식은 PMI의 개념을 활용한다는 점에서 다른 연구들의 방식과 같으나, 문맥정보를 반영하여 계산할 수 있도록 변형하여 표현하였다.

$$PMI(\text{표현어휘}, \text{상품특징}) \tag{4}$$
$$= \log_2 \frac{p(\text{표현어휘}, \text{상품특징})}{p(\text{표현어휘})p(\text{상품특징})}$$

$$Sentiment-Orientation(\text{표현어휘}, \text{상품특징}) \tag{5}$$
$$= pPMI(\text{표현어휘}, \text{상품특징}) - nPMI(\text{표현어휘}, \text{상품특징})$$
$$= \log_2 \frac{pP(\text{표현어휘}, \text{상품특징})}{pP(\text{표현어휘})pP(\text{상품특징})} - \log_2 \frac{nP(\text{표현어휘}, \text{상품특징})}{nP(\text{표현어휘})nP(\text{상품특징})}$$

(4)번 식은 (1)번 식에서 표현어휘와 상품특징어휘로 대입시킨 것이며, (5)번 식은 (2)번식과 마찬가지로 긍정적의미의 PMI합과 부정적의미의 PMI합의 차이 값으로 최종 극성을 구하는 것이다. 하지만 기존 방식과 달리, (5)번식은 문맥적 상황정보를 고려하여 계산할 수 있도록 변형시킨 것이며, 본 연구에서 따로 구축한 긍정적 또는 부정적 말뭉치사전을 활용한 계산방법이다. $pP(word)$값은 상품리뷰로부터 추출한 긍정적의미의 어휘들만을 포함하고 있는 사전을 대상으로 해당 어휘가 출현한 리뷰문서의 확률을 의미하며, $nP(word)$값은 부정적의미의 어휘들만을 포함하고 있는 사전을 대상으로 한 확률 값을 의미한다. $pP(word1, word2)$값과 $nP(word1, word2)$값의 경우에는 두 어휘가 같은 문장 내에 나타나는 경우의 리뷰문서의 확률을 의미한다. 이와 같이, 두 어휘의 긍정적인 관계 값을 의미하는 $pPMI$와 부정적인 관계를 의미하는 $nPMI$는 긍정적, 부정적 의미의 어휘들의 집합인 두 말뭉치사전을 활용한 PMI값 계산에 의하여 구해진다. 의미극성 판단의 중요한 역할을 하는 긍정적, 부정적 말뭉치사전에 대해서는 다음절에서 자세히 다루기로 한다.

4. 말뭉치사전의 자동구축

4.1 말뭉치사전 자동구축의 필요성

앞에서 언급한 바와 같이 의견분류의 성능에 있어서 무엇보다 중요한 것은 PMI값 계산 시, 비교 대상이 되는 말뭉치사전이다. 기존의 연구방식들에서는 이 말뭉치사전을 전문가가 미리 수작업으로 긍정적, 부정적 어휘를 몇 가지 정의하고, 워드넷 또는 웹문서를 대상으로 상호출현빈도를 계산하는 방법을 취하고 있다. 하지만, 이러한 경우 너무 일반적인 성향의 어휘의미를 따르고, 수작업에 대한 부하가 생긴다는 점에서 문제가 발생한다. 따라서 본 연구에서는 어휘가 사용된 상황적 문맥정보를 반영하기 위한 말뭉치사전 구축 전략 및 자동 구축을 통한 전문가 수작업 업무 제거에 대하여 다루었다. 이와 관련하여 자동 구축에 대한 타당성을 실험을 통해 밝히고, 어휘의 상황적 문맥정보를 반영한 말뭉치사전 구축을 위해 상품리뷰를 사용하는 것에 대한 타당성을 고찰하였다.

4.2 리뷰와 리뷰점수의 상관성

본 연구에서는 상품리뷰 요약에서의 의견분류에 사용되는 말뭉치사전 구축을 위하여 리뷰본문의 내용을 활

표 1 리뷰점수에 따른 리뷰제목 평가

리뷰점수	표본 수	좋음	혼재	나쁨	상품명,단순서술
5	217	129	5	1	82
4	216	115	8	1	92
3	216	82	42	22	70
2	216	19	16	97	84
1	215	5	9	116	85
총계	1,080	350	80	237	413

용하였다. 리뷰본문의 활용 가능성을 파악하기 위하여 통계적 실험을 수행하였으며 다음과 같은 자료 및 분석을 통해 리뷰본문을 말뭉치사전에 활용할 수 있음을 알 수 있었다.

실험에 사용된 대상 데이터는 ePinions.com의 휴대전화와 디지털카메라분류에서 1,080건의 데이터를 무작위로 선정하였으며 리뷰내용에 대한 자세한 분포는 표 1과 같다. 리뷰들에 대한 의미 판단은 20명의 연구원들이 수작업으로 관찰 및 판단한 것을 기초로 하였다.

표 1의 가로축은 리뷰제목의 실제 내용이 상품에 대하여 어떤 평가를 내리고 있는가를 나타내며, '상품명, 단순서술'의 경우는 어떠한 의견이 없이 단순히 서술한 경우를 의미한다. 표 1의 내용을 통해 전체에서 약 62%의 리뷰제목이 상품에 대한 의견을 표현하고 있는 것을 알 수 있다. 즉, 전체 리뷰 중, 62% 이상의 리뷰는 긍정적 또는 부정적인 의견을 파악할 수 있는 대상이 된다는 것을 알 수 있다. 의견을 표현하고 있는 리뷰를 말뭉치사전 구축에 활용한다고 했을 때, 그 내용의 구성을 파악하기 위하여 리뷰점수에 따른 평가내용의 분포를 살펴보았다. 그림 1에서 보는 것과 같이 높은 리뷰점수가 부여된 리뷰에서의 리뷰내용은 좋은 평가가 압도적으로 많았고, 반대로 낮은 리뷰점수가 부여된 리뷰에서의 리뷰내용은 나쁜 평가가 많았다. 중간 수준의 리뷰점수를 가진 리뷰의 경우 상대적으로 여러 의견들이 혼재하고 있는 것을 알 수 있다. 이를 통하여 직관적으로 리뷰본문과 실제 리뷰의 평가 사이에는 상관관계가 있음을 알 수 있다.

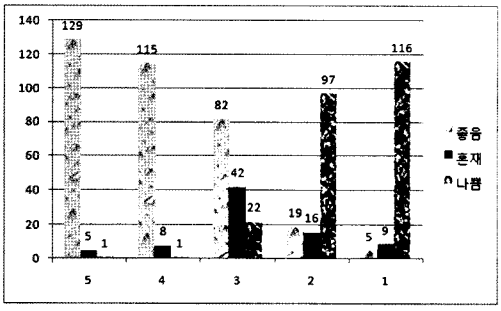

그림 1 리뷰점수에 따른 평가 분포

4.3 리뷰점수와 평가의 상관관계 분석

본 절에서는 리뷰점수와 리뷰에서의 실제 평가가 가지는 상관관계에 대하여 분석하였다. 4.2에서 제시한 데이터를 보면 직관적으로 두 요소사이의 상관성이 상당히 높은 것을 알 수 있지만, 상관도의 객관성을 더하기 위하여 피어슨 상관관계분석 수행하였고 그를 위한 가설은 다음과 같이 정의하였다.

귀무가설 (H_0): 리뷰점수와 실제 평가 사이에는 상관관계가 존재하지 않는다.

대립가설 (H_1): 리뷰점수와 실제 평가 사이에는 상관관계가 존재한다.

분석을 위해 평가정도에 따라 좋음=3, 혼재=2, 나쁨=1과 같이 값으로 표현한 후, 앞서 밝힌 표본들에 대하여 계산된 기술통계량은 표 2와 같다. 또한 상관계수정보는 표 3에서 보는 바와 같다. 두 변수의 유의확률이 0.01보다 작으므로 귀무가설은 성립되지 않고, 대립가설이 성립함을 알 수 있고, 리뷰점수와 실제 평가는 99%의 유의수준에서 0.76만큼의 상관도를 갖는 것을 알 수 있다. 즉, 두 변수 사이에는 상당한 연관성이 있으며, 리뷰의 내용이 해당 리뷰의 점수와 같은 성향을 띠고 있다고 말할 수 있다.

표 2 기술 통계량

	표본수	최소값	최대값	평균	표준편차
리뷰점수	667	1.00	5.00	3.02	1.4
평가	667	1.00	3.00	2.19	0.92

표 3 리뷰점수와 평가의 상관관계

		리뷰점수	평가
리뷰점수	피어슨 상관계수	1	0.76
	유의확률		0.00
	사례 수	667	667
평가	피어슨 상관계수	0.76	1
	유의확률	0.00	
	사례 수	667	667

4.4 리뷰본문을 활용한 말뭉치사전 자동구축

우리는 앞에서 언급된 내용을 기반으로 어휘의 문맥적 상황정보를 반영하기 위하여 해당 상품군의 리뷰제목 정보를 활용할 수 있음을 알 수 있었다. 약 62%의 리뷰제목에서 상품에 대한 명시적인 의견들이 관찰되었으며 상관관계분석을 통하여 리뷰내용이 리뷰점수와 같은 성향을 따른다는 것을 밝힘으로써 리뷰의 긍정적, 부정적 판단이 가능하게 되었다.

리뷰본문으로부터 말뭉치사전을 구축하기 위해서, 리

뷰본문을 대상으로 POS 태깅을 수행하였다. POS 태그 중에서 의견을 표현하는 어휘 및 특징을 나타내는 어휘가 될 수 있는 후보 태그들을 정의하여 해당 어휘들만을 대상으로 말뭉치 사전을 구축하였다. 말뭉치 사전은 의견분류에서의 활용을 위하여 긍정적 표현어휘사전, 부정적 표현어휘사전, 긍정적 특징어휘사전, 부정적 특징어휘사전 등 총 4가지로 구성된다. 말뭉치 사전의 구축에 사용된 POS 태그정보는 표 4와 같다.

표 4 말뭉치사전 구축 판단 기준

표현어휘 선정기준	POS tag가 'VBN(과거분사)', 'RB(부사)', 'RBS(최상급 부사)', 'JJ(형용사)', 'JJS(최상급형용사)', 'JJR(비교급 형용사)'인 단어
특징어휘 선정기준	POS tag가 'NN(명사)', 'NNP(고유명사)', 'NNS(복수형명사)', 'NNPS(복수형고유명사)'인 단어
긍정적 의미	리뷰점수가 5, 4점인 경우
부정적 의미	리뷰점수가 2, 1점인 경우

5. 실 험

5.1 실험 설계

본 장에서는 앞에서 언급한 상황인지 의견분류 기법의 성능을 평가하기 위한 실험을 소개하고 실험 결과에 대한 고찰을 서술한다.

실험데이터는 ePinions.com의 실제 상품리뷰데이터를 사용하였으며, 자세한 구성은 표 5와 같다. 휴대전화분류에 대한 상품리뷰데이터 2,947건에 대하여 2,196건의 긍정적 리뷰, 418건의 부정적 리뷰를 구분하였으며, 상품특징 48개를 선정하였다. 의견정보를 파악하기 위하여 선정된 48개의 상품특징들을 의견표현 어휘들과의 쌍으로 구성하였으며, 총 3,136쌍의 의견정보를 추출하였다. 또한, 디지털카메라분류의 12,917개 상품리뷰를 대상으로 37개의 상품특징을 선별하였으며 3,476쌍의 의견정보를 추출하였다. 기본적으로 의견정보 추출과정은 POS태깅정보 및 어휘간 거리, 문장 구성 등의 정보를 활용하였기 때문에, 실제 상품특성을 수식하고 있는 정보를 찾기 위한 정제작업이 필요하다. 따라서, 수작업을 통해 두 데이터셋의 의견정보쌍을 각각 734, 974 쌍으로 줄이는 작업을 수행하였으며, 'good', 'perfect', 'poor' 등과 같이 문맥에 영향을 받지 않고 고정된 의미성향을 갖는 표현어휘와 'big', 'small', 'high' 등과 같이 문맥에 따라 다른 의미성향을 가질 수 있는 표현어휘 그룹으로 나누었다. 이와 같은 의견정보의 구성을 표 5에 나타내었다.

표 5에 나타난 것과 같이 상품리뷰에 나타난 여러 의견정보 중에는 의견을 표현한 어휘만의 의미판단으로는 정확한 의미를 판별할 수 없는 경우가 다수 존재한다. 예를 들어 'big'라는 어휘가 'LCD'를 수식하는 경우와 'Body'를 수식하는 경우는 그 의견의 극성이 다르게 된다. 또한 같은 <표현어휘, 상품특징>의 쌍이라 해도, 사용자가 어떤 평가를 내렸는가에 따라 그 의미하는 바가 다르게 된다. 의미정보를 구성하는 의미표현 어휘들 중에는 'nice', 'wonderful' 등과 같이 그 성향이 명확한 어휘들이 있는 반면, 'big', 'small', 'many', 'few' 등과 같이 경우에 따라 그 의미의 극성이 확연히 달라지는 경우도 약 15% 내외 존재하는 것을 알 수 있었다. 본 실험에서 활용한 문맥에 따라 다른 의미성향을 갖는 어휘는 다음과 같다.

표 5 상품리뷰에서의 의견정보 분포

	휴대전화	디지털카메라
리뷰 수	2,947	12,917
긍정적 리뷰 수	2,196	9,940
부정적 리뷰 수	418	1,740
상품특징 수	48	37
추출된 의견정보 수	3,136 쌍	3,476 쌍
선별된 의견정보 수	734 쌍 (100%)	974 쌍 (100%)
고정된 의미성향을 갖는 표현어휘	610 (83.1%)	837 (85.9%)
문맥에 따라 다른 의미성향을 갖는 표현어휘	124 (16.9%)	137 (14.1%)

문맥에 따라 다른 의미성향을 갖는 어휘	big, small, high, low, long, short, many, much, few, several, huge, tiny, heavy, light, loud, silent

본 실험의 주요목적은 의견표현 어휘의 성향을 판단할 때 어휘의 문맥정보를 고려함으로써 얻을 수 있는 장점을 밝히는 것이다. 즉, 특정 어휘가 사용된 상황에 따라서 긍정적 또는 부정적 의미를 가질 수 있으므로 상황적 의미에 가장 알맞게 어휘의 의미적 성격을 판별하는 것을 목적으로 한다. 이를 위하여 다양한 상품특징들과 함께 나타나며 경우에 따라 그 의미가 달라질 수 있는 표현어휘들을 중심으로 실험을 수행하였다. 비교대상이 되는 방법은 2.2절에서 언급한 PMI를 이용한 기존연구방법을 활용하였다. 앞에서 설명한 것과 같이 기존연구방법 및 본 연구에서 제시한 방법 모두 PMI값을 활용하여 특정 어휘의 의미적 성향을 판별하게 되는데 계산된 PMI값의 절대치는 의미의 강도를 나타낸다. 하지만 본 실험에서는 올바른 의미적 극성을 판별해 내는 것을 평가하였다.

5.2 문맥고려 의견분류 결과

앞에서 언급한 약 1,700여건의 <상품특징, 표현어휘> 쌍에 대하여 본 연구에서 제시한 의견분류기법을 적용

한 결과와 기존 방식과의 비교실험 결과를 표 6에 나타내었다. 본 실험에서 비교대상으로 한 방법은 식 (2),(3)과 같은 방식을 사용한 Turney의 방법이다[3,4]. 정확성을 판단하기 위한 기준으로 본 연구에서 활용한 것은, ePinions.com의 리뷰데이터 중에서 'Pros', 'Cons' 데이터를 활용하였다. 두 데이터는 보통의 리뷰데이터들과는 다른 ePinions.com이 가지고 있는 특별한 데이터이며, 리뷰본문과는 별도로 각 상품리뷰별로 상품에 대해 장점 및 단점을 요약 및 구분해 놓은 것이다. 각 <상품특징, 표현어휘> 쌍이 장점 및 단점 데이터에 출현하는 정보를 기반으로 실험을 위한 평가기준을 만들었다. 또한, 평가기준의 정확성을 위하여 연구원들의 수작업을 통한 검증절차를 거쳤다.

이러한 평가기준을 가지고 본 연구에서 제시한 문맥정보를 고려한 의견분류 방식과 기존방법의 의견분류 정확도를 비교하여 아래의 표에 나타내었다.

표 6 의견분류 결과 정확도 비교

(%)	(가)		(나)		(다)	
	기존 방법	문맥 고려	기존 방법	문맥 고려	기존 방법	문맥 고려
휴대 전화	83.4	77.5	50.8	84.7	78.6	78.4
디지털 카메라	86.6	75.8	51.5	79.7	81.7	76.4
(가) 고정된 의미를 가진 어휘 (나) 문맥에 따라 다른 의미를 갖는 어휘 (다) 모든 어휘						

위의 표에서 보는 것과 같이 모든 어휘에 대한 경우 기존방법이 약간 좋은 성능을 나타냄을 알 수 있었다. 하지만, 앞에서 말한 어떤 상품분류 및 상품특징을 수식하느냐에 따라 의미가 달라지는 어휘, 즉 문맥에 따라 다른 의미성향을 갖는 어휘의 경우 본 연구에서 제시한 방식이 월등한 성능을 보임을 알 수 있었다.

실험결과를 분석한 결과, 고정된 의미를 가진 어휘에 대해서는 기존방법이 약간 높은 성능을 보임을 알 수 있었으며, 이것은 기존방법의 경우 대량의 웹문서를 대상으로 하여 어휘의 일반적의미를 활용한 분석방법이기 때문이다. 한편, 본 연구에서 제시한 방법의 경우, 어휘가 사용된 문맥정보를 활용함으로써 문맥에 영향을 많이 받는 어휘에 대하여 정확도가 높았다. 이러한 실험결과를 바탕으로 본 연구에서는, 두 방법의 장점을 모두 취하고 단점을 서로 보완할 수 있도록 두 방식에서의 PMI 점수를 조합하여 평균점수로 최종 PMI값을 계산하는 실험을 수행하였다. 두 방법 모두 어휘의 의미적 성향을 $\log_2$ 규모의 값으로 표현하고 있으므로 두 값의 평균을 최종 의미 분류에 활용하였다. 그 결과는 표 7과 같다.

표 7 조합된 방식의 의견분류 결과 정확도

	휴대전화	디지털카메라
모든 어휘	84.47%	85.7%
고정된 의미성향을 갖는 어휘	84.75%	88%
문맥에 따라 다른 의미성향을 갖는 어휘	83.06%	71.7%

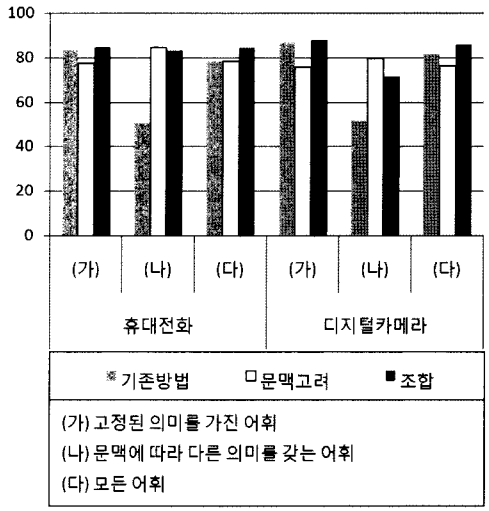

그림 2 의견분류방법 정확도 비교

그림 2에서 보는 것과 같이, 조합된 방식의 결과가 모든 경우에서 높은 정확도를 보임을 알 수 있다. 이러한 결과는 문맥에 따라 다른 의미를 갖는 어휘에 대한 의미성향 분류에서 본 연구에서 제시한 문맥을 고려한 의견분류방법이 높은 정확도를 가짐으로 인하여 조합된 최종 분류결과의 정확도 상승에 기여했기 때문이다. 또한, 두 방법에서 서로 상반된 극성의 결과를 분류해 낸 경우 계산되어진 PMI값의 평균을 통한 판단과정을 거침으로써 좀 더 높은 정확도를 보일 수 있었다.

표 8에는 문맥정보를 고려한 의견분류의 대표적인 예들을 나타내었다. 표의 (1),(2),(21) 등과 같이 표현어휘 자체의 의미가 명확하게 구분되어 지는 경우 해당 의미 극성을 잘 분류하였다. 본 연구의 특징으로 언급한 바와 같이 'big', 'small', 'high', 'low', 'long', 'short', 'many', 'few' 등과 같이 문맥적 상황에 따라 의미가 달라지는 표현어휘에 대해서는 어떤 상품분류의 어떤 상품특징을 서술하는가에 따라서 그 의미극성이 모두 다름을 예를 통해 알 수 있다. 또한, (4), (6) 등과 같이 상품특성의 성질이나 특징을 서술하는 표현어휘의 경우에도 그 의미 극성을 판별해 낼 수 있다.

5.3 논의

앞에서 언급한 것과 같이 본 실험을 통해서 밝히고자

표 8 문맥정보를 고려한 의견분류 예

번호	상품분류	상품특징	의견표현 어휘	의미극성
(1)	디지털 카메라	battery	good	긍정적
(2)		battery	poor	부정적
(3)		battery	short	부정적
(4)		image	red-eye	부정적
(5)		image	tiny	부정적
(6)		image	film-like	긍정적
(7)		lcd	big	긍정적
(8)		lcd	swivel	긍정적
(9)		memory	big	긍정적
(10)		price	high	부정적
(11)		price	low	긍정적
(12)		quality	high	긍정적
(13)		quality	low	부정적
(14)		resolution	high	긍정적
(15)		resolution	low	부정적
(16)		video	long	긍정적
(17)		zoom	high	긍정적
(18)		zoom	low	부정적
(19)	휴대전화	button	large	긍정적
(20)		call	long	긍정적
(21)		color	vivid	긍정적
(22)		design	intuitive	긍정적
(23)		display	dual	긍정적
(24)		feature	few	부정적
(25)		feature	many	긍정적
(26)		problem	many	부정적
(27)		problem	few	긍정적
(28)		function	many	긍정적
(29)		ring	loud	긍정적
(30)		ring	small	부정적
(31)		weight	low	긍정적

한 주요 내용은 의견을 표현하고 있는 어휘들의 문맥적 상황정보를 고려한 의견분류를 통해서 상품리뷰에 나타난 의견이 갖는 의미적 극성을 정확하게 판별하고자 한 것이다. 실험 결과에서 알 수 있듯이 'big', 'small', 'many', 'few' 등의 의견을 표현하는 어휘들은 그 자체의 의미보다는 어떤 상품의 어떤 특징에 대해서 서술되었는가가 고려되어야 정확한 의미를 판단할 수 있다.

본 연구에서는 어휘가 사용되는 문맥정보를 정의하고 의견분류에 적용하였기 때문에 상황별 어휘의 의견 성향을 정확히 추정할 수 있었다. 이것은 리뷰점수를 활용하여 말뭉치사전을 자동으로 구축하여 사용하는 한편, <표현어휘, 상품특징>의 쌍으로 의견정보를 처리함으로써 어휘의 문맥정보를 모두 반영함으로써 얻을 수 있는 결과이다.

기존 의견분류 방법과 본 연구의 또 하나의 차이점은 실제 상품리뷰 분석 환경에서의 적용 가능성에 있다. 수작업으로 말뭉치를 정의하고 의견분류를 수행하는 방법의 경우 수천 또는 수만 개의 상품에 대한 리뷰들을 모두 다루기 힘들다. 하지만, 본 연구에서 제시한 방법으로는 상품리뷰로부터 말뭉치를 자동으로 구축하고, 이를 의견분류에 활용함으로써 많은 상품리뷰 데이터를 빠르고 쉽게 분석할 수 있다.

실험을 통하여 문맥에 따라 그 의견 성향이 영향을 받는 어휘의 경우 본 연구에서 제시한 방법의 우수성을 밝혔으며, 기존 방법의 장점도 이용하기 위하여 두 방법의 조합된 방식에 대한 정확성 또한 밝혔다. 따라서, 두 방법의 조합을 통해 기존 방법의 문맥에 따른 어휘 의견분류의 취약성을 극복하는 것이 최상의 결과를 보인다는 것을 알 수 있었다.

6. 결론 및 향후과제

본 논문에서는 어휘의 의견분류를 수행할 때, 기존방식들이 어휘의 일반적 의미를 활용하고, 많은 수작업이 요구된다는 단점을 보완하는 의견분류에 대한 방법을 제시하였다. 이를 위하여 상품리뷰 요약에서의 의견표현 어휘가 갖는 문맥정보를 정의하고, 각 문맥정보를 어떻게 획득하여 적용시키는가에 대한 방법을 논하였다. 또한, 의견분류과정에서 사용되는 말뭉치사전을 자동으로 구축하는 방법도 제시하였다. 마지막으로 실험을 통하여 특정 상품군에서 각 상품특징에 따른 의견표현어휘의 의미 판별 정확성을 밝혔다. 또한 본 연구에서 제시한 방법을 활용하여 기존방법의 취약성을 극복할 수 있음을 실험을 통하여 밝혔다.

향후 과제로서 본 연구에서 제시한 통계적 방법에 자연어처리 방법을 접목하여 의견분류의 정확도를 더욱 높이는 연구를 진행하고 있다. 또한 구어체로 표현된 상품리뷰의 내용을 추출하여 의견분류에 활용하는 방법 또한 필요할 것이다.

참 고 문 헌

[1] M. Hu, B. Liu, "Mining and summarizing customer reviews," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp.168-177, 2004.

[2] D. Lee, O. Jeong, S. Lee, "Opinion Mining of Customer Feedback Data on the Web," *Proceedings of The Second International Conference on Ubiquitous Information Management and Communication*, pp.230-235, 2008.

[3] X. Ding, B. Liu, P. S Yu, "A holistic lexicon-based approach to opinion mining," *Proceedings of the international conference on Web search and web data mining*, pp.231-240, 2008.

[4] Turney, P. D., "Thumbs up or thumbs down?:

semantic orientation applied to unsupervised classification of reviews," In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, 2002.

[5] Turney, P. D., Littman, M. L., "Measuring praise and criticism: Inference of semantic orientation from association," *ACM Trans. Inf. Syst.*, **21**, 4, Oct. 2003.

[6] Namrata Godbole, Manjunath Srinivasaiah, Steven Skiena, "Large-Scale Sentiment Analysis for News and Blogs," Int'l AAAI Conference on Weblogs and Social Media (ICWSM 2007), 2007.

[7] E. Boiy, P. Hens, K. Deschacht, M. Moens, "Automatic Sentiment Analysis in On-line Text," *ELPUB2007 Conference on Electronic Publishing*, June 2007.

[8] J. Yi, W. Niblack, "Sentiment Mining in WebFountain," *International Conference on Data Engineering (ICDE'05)*, pp.1073-1083, 2005.

[9] T. Nasukawa, J. Yi, "Sentiment analysis: capturing favorability using natural language processing," *Proceedings of the K-CAP-03, 2nd International Conference on Knowledge Capture*, pp.70-77, 2003.

[10] Xiaowen Ding, Bing Liu. "The Utility of Linguistic Rules in Opinion Mining," pp.811-812, SIGIR2007, 2007.

[11] M. Gamon, A. Aue, S. Corston-Oliver, E. Ringger. "Pulse: Mining Customer Opinions from Free Text," In *Lecture Notes in Computer Science*, Vol.3646. Springer Verlag. (IDA 2005), 2005.

[12] Wilson, T., Wiebe, J., Hoffmann, P. "Recognizing contextual polarity in phrase-level sentiment analysis," In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pp.347-354, 2005.

[13] Scaffidi, C., Bierhoff, K., Chang, E., Felker, M., Ng, H., Jin, C. "Red Opal: product-feature scoring from reviews," In *Proceedings of the 8th ACM Conference on Electronic Commerce 2007*, pp.182-191, 2007.

[14] Miller, G., Beckwith, R, Fellbaum, C., Gross, D., Miler, K. "Introduction to WordNet: An on-line lexical database," International Journal of Lexicography, pp.235-244, 1990.

양 정 연

2002년 충남대학교 컴퓨터공학과 학사 2003년~현재 서울대학교 컴퓨터공학부 석박통합과정 재학중. e-Business 기술 연구센터 연구원. 관심분야는 Opinion Mining, e-Business Technology, Databases, Semantic Technology

명 재 석

2007년 성균관대학교 정보통신공학부 학사. 2007년~현재 서울대학교 컴퓨터공학부 석박통합과정 재학중, e-Business 기술 연구센터 연구원. 관심분야는 Opinion Mining, e-Business Technology, Databases, Semantic Technology

이 상 구

1985년 서울대학교 계산 통계학과 학사 1987년 M.S. Computer Science, Northwestern University, Evanston, Illinois 1990년 Ph.D. Computer Science, Northwestern University, Evanston, Illinois ~현재 서울대 컴퓨터공학부 교수, e-Business 기술 연구 센터장. 관심분야는 e-Business Technology, Databases, Opinion Mining, Mobile Database